Invention Journal of Research Technology in Engineering & Management (IJRTEM)
ISSN: 2455-3689
www.ijrtem.com Volume 2 Issue 8 ǁ August 2018 ǁ PP 01-03

Data Warehouse: A Primer

[1,]Matthew N. O. Sadiku, [2,]Chandra M. M. Kotteti, and [3,]Sarhan M. Musa
Roy G. Perry College of Engineering Prairie View A&M University Prairie View, TX 77446

ABSTRACT: *Data warehousing is a technique for collecting and managing data from multiple internal and external sources to provide meaningful business insights. Data warehouses are designed to give a long-range view of data over time and provide a decision support system environment. They are a vital component of business intelligence, which is designed for data analysis and reporting. They are used to provide greater insight into the performance of a business. This paper provides a brief introduction on data warehousing.*

KEY WORDS*: data warehouse, enterprise data warehouse, data repository, business intelligence*

I. INTRODUCTION

The problem of generating, storing, and analysing immense volumes of data has led to the need of large data centralized repositories, e.g. data warehouses. Data warehouses (DW) are data repositories specialized in supporting decision making. They are typically used to correlate broad business data to provide greater executive insight. The term "data warehousing" was introduced in 1988 by IBM researchers Barry Devlin and Paul Murphy. Later in 1990, Bill Inmon (regarded as the father of data warehousing) defined DW as a subject-oriented, integrated, time-invariant, and non-volatile collection of data. DW is a relational database that is designed for query and analysis rather than for transaction processing. The relational or decision support database (data warehouse) is maintained separately from the business' operational database [1]. Operational data supports the day-to-day operations of the business, while information data is required for making sound, strategic business decisions. A company might warehouse data for use in exploration, looking for patterns of information that will help them improve their business processes. DW is regarded as a core component of business intelligence. In that regard, it is a technology that aggregates structured data from one or more sources for comparison and analysis for greater business intelligence. There are different ways of storing data in a data warehouse. The most important approaches are the dimensional approach and the normalized approach.

DW CHARACTERISTICS

As mentioned earlier, Bill Inmon defined DW as a collection of data that exhibits the following four characteristics.

- ***S****ubject-oriented*: A DW is subject-oriented since it provides information around a subject, which may be products, customers, sales, etc. All the data items related to the same subject are connected.
- *Integrated*: Data may need integration since they may be distributed across heterogeneous sources such as other relational databases, flat files, etc. The data integration is based on a model of the enterprise.
- *Time-invariant*: Data has element of time in that it is identified with a particular time period. The data is relevant to some moment in time.
- *Non-volatile:* Data is not subject to frequent modification. The previous data is not erased when the new data is added to it. Data are read-only and are never updated or deleted.

II. HOW DATA WAREHOUSE WORKS

A data warehouse merges information coming from different sources (finance, billing, personnel, etc.) into one comprehensive database. Data flows into a data warehouse from the transactional system, other relational databases, or operational systems. The data is extracted, cleansed, transformed, and loaded in the data warehouse. The data is stored as a series of snapshots, where each record represents data at a specific time. The data may be structured, semi-structured, or unstructured [2]. Data brought into a DW must have the same format and have the same units. The main idea of data warehousing is separating OLAP (On-Line Analytical Processing) queries from OLTP (On-Line Transactional Processing) queries. OLAP is characterized by a relatively low volume of transactions, while OLTP is characterized by a large number of short on-line transactions. Traditional online transaction processing (OLTP) databases automate day-to-day transactional operations. OLTP databases are optimized for data storage and strive to eliminate data duplication. Data warehouse can be implemented in several different ways leading to different types of DW such as enterprise

data warehouse, operational data store, and data mart. A database may be oriented towards one or more business areas, such as sales, tracking inventories or transactions [3]. Several system architectures can be developed for a DW. A typical three-tier architecture is shown in Figure 1 [4].

APPLICATIONS

Typical applications of DW are data-intensive; they involve huge amounts of data that push database management technology to the limit. DW is commonly applied in healthcare, data mining, and higher education.

- *Healthcare*: There has been an ever increasing volume of data being collected in the healthcare industry. Healthcare DW includes various data such as hospital discharges, bills, births, deaths, and disease registries. It supports decision making activities and health assessment of communities [5].
- *Data mining*: Data mining has become an essential element of a computer strategy. Data warehousing makes data mining possible. Data mining is looking for patterns in the data that may lead to higher sales and profits.
- *Higher education*: The academic data of higher institutions is growing significantly. It is essential for colleges and universities to adopt DW to enhance their decision making [6].
  DW is also used in census analysis, financial services, manufacturing, banking, airline, telecommunication, hotel industry, retail chain, insurance, transportation, and government.

BENEFITS AND CHALLENGES

The most popular benefit of data warehousing is simplicity; it simplifies decision making. Other benefits include [7].

- Reduced effort by developers to produce information
- Improved user ability to produce information
- More and better information
- Increased decision-making speed
- Improvement for business processes
- High query performance
- High returns on investment
- Support for the accomplishment of strategic business objectives
- Increased productivity and efficiency
- Substantial competitive advantage

There are a number of challenges still in implementing DW. These include [8]:

- *Scalability*: It is often difficult to anticipate the number of users accessing the DW simultaneously.
- *Speed*: Several factors (distance, devices, interference, noise) contribute to the transfer speed on the Internet.
- *Security*: Since its advent, DW has gone through many changes which have caused changes in the security strategies as well. Network security involves either unauthorized access to private data or computer viruses corrupting data.

Although the DW process supports bottom-up extraction of data, it fails in top-down enforcing the business strategy. The ability of some companies to manipulate data lags behind the growth rate of the data.

III. CONCLUSION

The need to data warehouse data evolved as computer systems became more complex and had to handle increasing amounts of data. Data Warehouses have steadily become important in organizations that possess large amount of data. Their implementation has revolutionized business, boosted customer service, and helped leaders make better decisions. They have become a key component in business intelligence. DW is converging with the Internet. When combined, the two technologies constitute enormous potential for corporations. Big data technology should be used to extend the existing DW solutions. This will help reduce the problem with traditional data analysis [9]. More information on data warehousing can be found in [10-17] and similar books available at Amazon.com.

REFERENCES

[1] "Data warehouse," Wikipedia, the free encyclopedia https://en.wikipedia.org/wiki/Data_warehouse

[2] "What is data warehousing? Types, definition & example," https://www.guru99.com/data-warehousing.html

[3] "Overview of data warehousing," https://www.ibm.com/support/knowledgecenter/en/SSGU8G_11.50.0/com.ibm.whse.doc/ids_ddi_344.htm

[4] K. Merritt, "Data warehouse and the Internet," Journal of Internet Commerce, vol. 1, no. 2, 2002, pp. 49-61.
[5] D. J. Berndt et al., "Healthcare data warehousing and quality assurance," Computer, 2001, pp. 56-65.
[6] I. M. Aljawarneh, "Design of a data warehouse model for decision support at higher education: A case study," Information Development, vol. 32, no. 5, 2016, pp. 1691-1706.
[7] H. J. Watson et al., "Current practices in data warehouse," Information Systems Management, Winter 2001, pp. 47-55.
[8] L. Chen and M. N. Frolick, "Web-based data warehousing: Fundamentals, challenges, and solutions," Information Systems Management, Spring 2000, pp. 80-86.
[9] L. W. Santoso and Yulia, "Data warehouse with big data technology for higher education," Procedia Computer Science, vol. 124, 2017, pp. 93-99.
[10] P. Ponniah, Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. John Wiley & Sons, 2001
[11] A. Berson and S. J. Smith, Data Warehousing, Data Mining, and Olap. New York, NY: McGraw-Hill, 1997
[12] B. Devlin and L. D. Cote, Data Warehouse: From Architecture to Implementation. Boston, MA: Addison-Wesley Longman, 1996.
[13] W. H. Inmon, J. A. Zachman, and J. G. Geiger, Data Warehousing and the Zachman Framework: Managing Enterprise Knowledge. New York, NY: McGraw-Hill, 1997.
[14] R. Kimball and M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. John Wiley and Sons, 3rd edition, 2013.
[15] W. H. Inmon, Building the Data Warehouse. John Wiley and Sons, 1996.
[16] S. Anahory and D. Murray, Data Warehouse in the Real World: A Practical Guide for Building Decision Support System. Addison-Wesley, 2002.
[17] C. Imhoff, N. Galemmo, and J. C. Geiger, Mastering Data Warehouse Design: Relational and Dimensional Techniques. Indianapolis, IN: Wiley Publishing, 2003.

ABOUT THE AUTHORS

Matthew N.O. Sadiku is a professor at Prairie View A&M University, Texas. He is the author of several books and papers. He is an IEEE fellow. His research interests include computational electromagnetics and computer networks.

Chandra M. M. Kotteti is currently a doctoral student at Prairie View A&M University, Texas. His research interests include fake news detection using machine learning and deep learning, natural language processing, big data analytics, and wireless networks.

Sarhan M. Musa is a professor in the Department of Engineering Technology at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow.

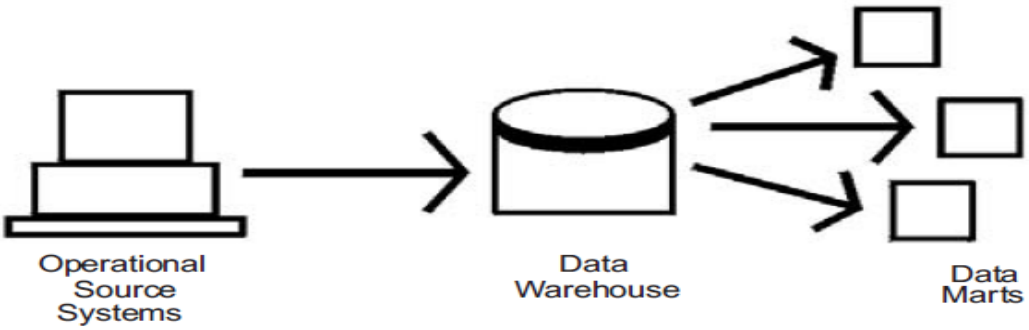

Figure 1 A three-tiered architecture for DW [4].

Matthew N. O. Sadiku (2018). Data Warehouse: A Primer. Invention Journal of Research Technology in Engineering & Management (IJRTEM), 2(8), 01-03. Retrieved August 8, 2018, from www.ijrtem.com.